**Doi:** 10.30473/IDEJ.2022.63913.1104 **ISSN:** 2588-4476

**Iranian Distance Education Journal**

Vol. 3, No. 2, (New Series) Summer-Autumn 2021 (P 32-44), Payame Noor University

**Original Article**

# Predicting students at risk of academic failure using learning analytics in the learning management system

**Hamid Zangooei*[1], Omid Fatemi[2]**

1. M.Sc. Student of Computer Architecture, Faculty of Engineering, Faculty of Electrical and Computer Engineering, University of Tehran, Iran.

2. Associate Professor, Department of Electrical and Computer Engineering, Campus of Technical Schools (Faculty of Electrical and Computer Engineering, University of Tehran), Tehran, Iran.

**Received:** 2021/06/22 **Accepted:** 2021/10/22

## Abstract

Online learning platforms have become commonplace in modern society today, but high dropout rates and decrement students' performance still require more attention in such online learning environments. The purpose of this research is to accelerate the identification of students at risk of academic failure in order to take appropriate corrective action. Therefore, we have proposed model to achieve this goal and ultimately improve the performance of students and faculty. Then, for early prediction of students at risk of academic failure, the short-term memory neural network (LSTM) and the widely used support vector algorithm have been used to analyze students' time based behaviors using data from the University of Tehran e-learning system. To demonstrate the optimal performance of the predictive algorithm, we compared the LSTM network with the support vector algorithm with different evaluation criteria. The results show that the use of LSTM network for early prediction of students at risk provides higher predictive accuracy compared to the support vector machine algorithm. In this research, our method in predicting students' performance with LSTM network has achieved 94% accuracy and with support vector machine algorithm has achieved 88% accuracy. In addition, the Area Under the Curve (AUC) was 0.936 and 0.882, respectively, using the LSTM algorithm and the support vector machine. Therefore, according to the obtained results, it can be seen that our proposed algorithm has an important and effective contribution to improving the final performance of teachers and students during the course.

## Keywords

Learning Analytics, Long Short Term Memory Network, Support Vector Machine, Predicting Students at Risk of Academic Failure

## Introduction

One of the major problems of the country's educational system is academic failure and the decrement of students' performance from a satisfactory level to an unfavorable level during the course. Most students do not have enough information about educational issues when entering the university and they encounter problems during their studies that need to be identified, prevented and decremented. Academic failure can lead to a lack of desire of students to continue their education and decrement their academic achievement. This issue also ultimately leads to irreparable damage to the services of the country's educational system. In addition, little attention is paid to the problem of students' academic failure due to the existence of mental and educational disorders and the cause of negative social behaviors such as dropout in students. If the basic steps are not taken to identify the relevant factors, prevent and decrement academic failure, the consequences will affect the country's education system. Therefore, it is necessary to study the

*Corresponding Author: H.zangooei@ut.ac.ir

factors related to academic failure in order to form an alert educational system that has the ability to identify and screen students at risk of academic failure. Because recognizing the factors related to academic failure and adopting preventive measures can help to decrement academic failure and it can improve the educational status of students.

Evaluating and analyzing data generated by students in the LMS system can assist faculty in understanding, reviewing and monitoring students' learning progress. Predicting and early prediction of students' performance in the LMS system, will lead to the provision of more appropriate solutions by teachers to prevent the failure in student's quality of education.

One of the important prerequisites for examining the educational status of students is the existence of sufficient data for this event. The speed of diagnosis and prediction of students' status is directly related to the probability of preventing their final academic failure. Therefore, predicting students' performance at the beginning of each semester will be more useful than predicting students' performance at the end of that semester. In addition, developing a customized forecasting model to identify students' learning behavior at the beginning of each semester is a difficult and challenging task.

In this research, the issue of academic failure, which is one of the main problems in the educational system of each country, has been studied. Using a series of data, we try to improve students' academic achievement. In fact, the main focus of this research is on how students should have upward growth during their studies and be informed if they suffer from academic failure. In addition, policies should be implemented that decisions can be made to prevent or eliminate these problems. We use students' data in the learning management system of the University of Tehran to predict the performance of their courses using the long short term memory neural network learning model. We trained the LSTM network with a set of data over different time periods to see how to achieve a very accurate prediction in the shortest amount of time. We also compared and evaluated the performance of the LSTM network compared to the SVM classifier in terms of predictive accuracy.

Before starting in order to clarify the scope of this research, we will ask some questions and during this research we will answer these questions.

-What parameters help to identify and predict students at risk of academic failure?
-What is the role of learning analytics tools in assessing students' performance during the course?
-What is the advantage of LSTM neural network compared to other classification in this research?
-What steps should be taken to learn a student's academic behavior?
-What are the strategies for receiving effectiveness from the results of this research?
-What people with what position and how can use the output of this research?
-What is the main purpose of this research?

The structure of this article is as follows. Section 2 summarizes past work. Section 3 presents the proposed method used. Section 4 is about analysis and evaluation. Section 5 shows the conclusion.

## Literature Review

Many studies of machine learning approaches to predict students' performance [1,2] and Improvement the learning environment [3] when student's behavioral data due to the development of learning environments such as MOOC and Learning Management System (LMS) in Available, used. Predicting student's outcomes during the semester in the MOOC and LMS is important because it can help teachers identify at-risk students and assist them in taking lessons. Multiple work has been done to identify at-risk students. Some previous research on students'performance prediction has used traditional machine learning methods to fit demographic information, interaction reports, or both.

Logistic regression was commonly used in models that predicted students at risk of failure, and showed promising predictive results. Wilson et al. [4] applied a logistic regression to participants' demographic information that matched their writing tasks and personal abilities, and the model produced a promising area under the curve (AUC) of 0.89.

Marbouti et al. [5] also used logistic regression to evaluate students' performance in academic achievement with face-to-face characteristics and to evaluate their behavior.

Aljohani et al. [6] compared logistic regression methods, support vector machines, simple Bayesian classification, and the J48 algorithm in predicting academic achievement / failure based on organizational data and tracking data generated by VLE, and The J48 algorithm has the best classification accuracy and the best execution time (with the exception of Naive Bayes). These machine learning methods show promising results in predicting students' performance with fixed length data.

Osmanbegovic et al. [7] analyze simple Bayesian (NB), decision tree (DT) and multilayer perceptron (MLP) algorithms to predict student's success. The data consists of two parts. The first part of the data is collected from a survey conducted at the University of Tuzla in 2010-2011. The participants were first year students in the Department of Economics. The second part of the data is obtained from the registration database. In total, the dataset has 257 samples with 12 features. They used Weka software as an implementation tool. Classifiers are evaluated using accuracy, learning time, and error rate. NB with a training time of less than 1 second and a high error rate achieves a high accuracy score of 76.65%. Baradwaj and Paul [8] also examine data mining approaches to predict students' performance. They check the accuracy of DT, where DT is used to extract valuable rules from the data set. The data set used in their review was obtained from the University of Purvarichal, India, which includes 50 student files, each with eight features.

Kovacic [9] analyzed the early prediction of students' success using machine learning techniques. This review examined various features such as education, work, gender, ethnicity, curriculum, etc. for effective prediction. This dataset was collected from the Free University of New Zealand. To select a feature, machine learning algorithms are used to identify the key features that affect students' success. The key findings were that ethnicity and curriculum are the two main features that affect students' success.

Watson et al. [10] considered the student's activity report in the introductory programming of a course to predict their performance. This research recommended a predictor based on automated measurement criteria rather than a direct basis for determining student's evolving performance over time. They proposed a scoring algorithm called WATWIN, which assigns specific scores to each student's programming activity. The scoring algorithm takes into account the student's ability to deal with programming errors and the time to resolve these errors. This review used the data of 45 student's programming activity report from 14 sessions as a data set. Each student's activity was assigned a WATWIN score, which is then used in linear regression. Linear regression achieves an accuracy of 76% using the WATWIN score. For effective forecasting, the data set must be balanced. Balanced data means that each prediction class has an equal number of features.

In this research, we incorporated the nature of time into the construction of our proposed features and created a new type of feature that was not used in previous researches. The values of these features change over time due to their dynamic nature. Previous researches had not considered the dynamic nature of students' performance during the academic period, but in this piece of research, by evaluating an optimal period model, we evaluated and modeled the evolutionary and changing performance of students. In addition, the use of LSTM and SVM neural networks with optimal parameters along with these cases further added to the quality of the model based on learning behavior. Therefore, this method has low computational complexity and overhead, and at the same time it is effective, efficient and with high accuracy.

## Proposed Solution

The solution of this research according to (Figure1) includes 6 steps. In the first step, data collection and refinement was performed. In the second step, labeling and data cleaning and time division were performed for data processing. In the third step, the features were extracted. In the fourth step, Pearson correlation was used to select the effective features. In the fifth step, the LSTM network and the support vector machine algorithm were used for early prediction. In the sixth step, the results obtained from the implementation of the algorithms as well as the comparison of the accuracy of the LSTM network and the support vector machine algorithm are discussed. Finally, we evaluate the General of the Approach.

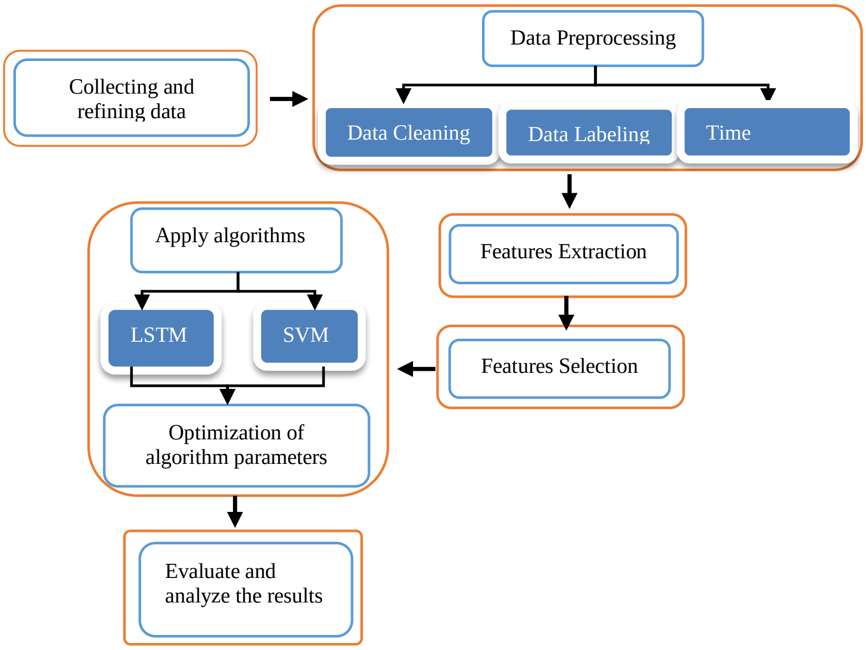


**Figure1.** General steps of our proposed method in the form of block diagrams

## 1. Collecting and Refining Data

In LMS, various information is recorded for each learner. According to the data mining objectives of this research, some of the features of the educational system have been considered that play a role in identifying the pattern of success and failure of learners.

The data set used in this research is the data collected from the learning management system of the University of Tehran to collect a rich data set and prepare it for testing. This data includes information about student's activities, exercises, grades, time spent, and more. The collection of this data is based on studies conducted on an educational system designed and implemented based on the structure of the Moodle e-learning system. This system provides all the facilities and operations related to the teacher and the student electronically.

With the Moodle e-learning system, teachers can offer their curriculum to students with more flexibility. Thus, students can benefit from their learning by accessing online activities, semester content, communication tools, and assessment by interacting more with the teacher. The entire data set is analyzed and a transparent body is extracted from it.

Twelve semester data were used to train, validate and test the prediction models. A total of 587 students participated in this course. We randomly used 80% of the data (469 students) for training and validation; the remaining 20% of the data (117 students) was used for testing.

## 2. Data Preprocessing

### 2.1. Data Cleaning

After collecting data through the LMS of the University of Tehran, each data was analyzed at first glance. Some items in the dataset were deleted at first glance. Since the focus of the research is on early prediction of students, in order to clean the data, items without values in the data set were removed first. Then, items with zero value and outliers were identified and deleted.

### 2.2. Data Labeling

In this section, the final grades are used as an indicator of students' performance. Two labels were used to balance the data. According to (Table1), if the student's final grade is in the range [0, 60], the student is shown with the label at risk and if the student's final grade is in the range [60,100], the student is shown with the label Pass.

**Table1.** Labeling

| Final grade | Label |
|---|---|
| 0%-60% | At risk |
| 60%-100% | Pass |

### 2.3. Time Division

The time division determines how many periods the information in the lesson is divided into, as well as the start and end of these periods. If a small period is taken, there will not be enough data to calculate, and if a large period is taken, the negative or positive effect of many events will be reduced due to the average of the properties taken. Therefore, according to (Figure2), the number of Periods was considered equal to 4, so that there is enough data to calculate.

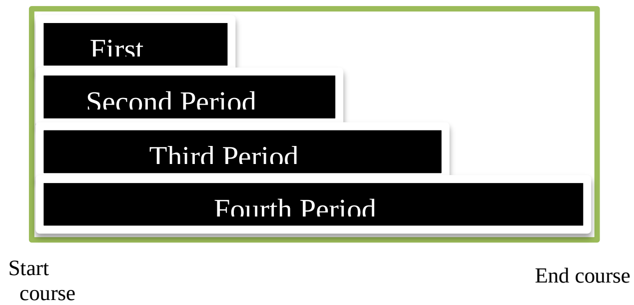


**Figure2.** Time Division

## 3. Features Extraction

In this section, various properties of the cleaned data are calculated and extracted. The purpose of feature extraction is to make raw data more usable for future statistical processing. For the early prediction of at-risk students, we suggest using time-based features. Time-based features include the amount of time a student spends doing quizzes, video feedback, and classroom activities. Students' activities in the LMS include doing tests, watching videos, doing video feedback, doing posts, doing workshops, and doing homework. By thoroughly examining the data set based on this system, 14 features were extracted. Which are specified according to (Table2).

**Table2.** Features Extraction

| Features | |
|---|---|
| Number of videos viewed | NVW |
| The score that the student got from the video | SSV |
| The number of quizzes has done | NQ |
| Quiz rate earned | RQ |
| Average scores obtained in quizzes | ASQ |
| Number of class work done (delivery) | NCD |
| Classwork rate earned (delivery) | RCD |
| Average grades obtained in the classwork (delivery) | ANCD |
| Number of class work done (evaluation) | NCE |
| Classwork rate earned (evaluation) | RCE |
| Average scores obtained in the classwork (evaluation) | ANCE |
| Number of feedbacks to the video | NFV |
| Feedback rate to the video earned | RFV |
| Average scores obtained in video feedback | ANFV |

## 4. Features Selection

Features with a correlation close to zero are being selected. Using Pearson method, linear correlation between both features was calculated. The result of these calculations for both numerical properties is between -1 and +1. The numbers +1, 0 and -1 indicate positive linear correlation, non-linear correlation and negative linear correlation, respectively. After calculating the correlation, it was observed that the correlation rate of the properties is not high, so the properties are independent. These features were then used in the student prediction model, which will be observed at the end, making full use of the feature set increases accuracy. According to (Figure3), features is performed in the form of Heat map analysis.

The correlation matrix in (Figure3) has 14 rows and 14 columns that are symmetric. The number of rows and columns is equal to the number of properties. Each of the houses is marked with a color that is in the range of +1 and -1. The closer this number is to -1, it means that the two properties (at the intersection of the two numbers) are inversely related to each other, and the closer this number is to +1, it means that the two properties are directly related to each other. The original diameter of this matrix is equal to +1 because each property naturally has a maximum correlation with itself.

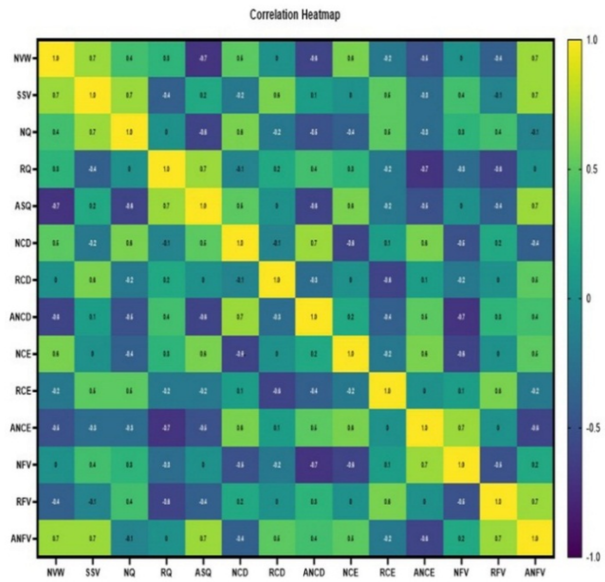


**Figure3.** Heat map

## 5. Realization of LSTM Network

LSTM network was used as a learning method for early prediction of at-risk students. The reason for choosing the LSTM network is that it remembers past information and, while predicting future values, also takes this past information into account. Also, the student's learning process is a sequence of time, which makes LSTM very suitable for time-dependent data. We proposed a customized model based on machine learning algorithm for the LMS of the University of Tehran.

The LSTM network is a subset of recurrent neural networks. As shown in (Figure4), loops allow information to be transferred from one time period to the next. Loops in LSTM look like a normal neural network when opened, but include multiple versions of the same neural network that transmit information from one network to the next. In particular, the output of a network in one time period is converted to its input in the next time period, which makes LSTMs very successful in modeling time series data [11].

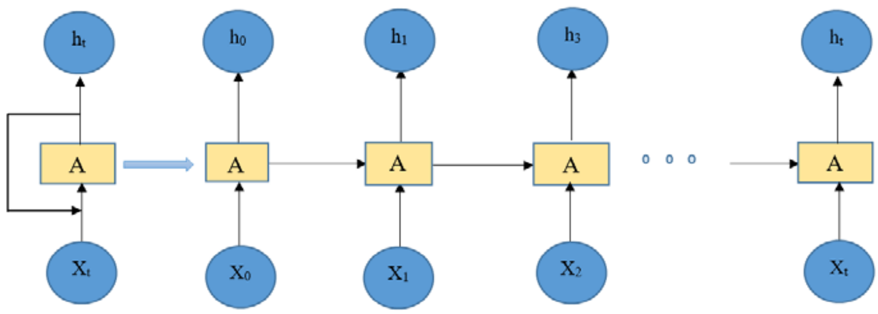


**Figure4.** Recurrent neural network

First, the features for each student in each time period were calculated, which resulted in the creation of an attribute vector with 4 time periods for the LSTM network. Then, according to (Figure4), for the early prediction in the first time period, the property vector X0 is taken as the input sequence and then h0 is delivered as the output, which together with X1 is the input of the next step. Therefore, h0 and X1 are the next step inputs. Also, for the early prediction of students, the second, third and fourth periods were implemented to show how well this model can have a good prediction in a short period of time.

Since the LSTM network consists of three gates, in a specific time instance such as t, the input gate is used to update the amount of embedded memory, the Ct memory cell is responsible for storing information. The output gate decides what the next hidden state is. Forget Gate (ft) is responsible for controlling information that is stored or forgotten. LSTM network gates are shown mathematically in Equations (a) to (f):

a) $f_t = \sigma (w_f [h_{t-1}, x_t] + b_f)$
b) $\xi_t = \mathrm{Tanh} (W_\xi [h_{t-1}, x_t] + b_\xi)$
c) $i_t = \sigma (W_i [h_{t-1}, x_t] + b_i)$
d) $C_t = f_t C_{t-1} + i_t \xi_t$
e) $o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$
f) $h_t = o_t * \mathrm{Tanh}(C_t)$

In Equations (a) to (f) bf, bξ, bi, bo, Wf, Wξ, Wi, Wo are the weights and biases of the gates and σ is a sigmoid function. Our research used a sequential model for the early prediction of students at risk based on interaction in periods. To achieve this goal, data sets were processed for each period. Features were recorded for each student in each period. As shown in (Figure5), where S1, S2… Sn show unique students that are the same for all periods.

**Figure5.** Architecture of the student's early prediction model

The vector consisting of sequences of periods for students was formed and transferred to the model sequentially. Therefore, in the LSTM network, it was possible to make early predictions for each period. The architecture of the student's early prediction model is shown in (Figure5).

We performed cross-validation to evaluate the students' early prediction model. Because the training and testing process is performed on several different parts of the data set, a more stable and reliable result is obtained. Twelve semester data were used to train and test the prediction model. A total of 587 students participated in this course. 80% of the data were randomly used for training; the remaining 20% of the sample was used for testing. In order to have a sufficient amount of data in each section, a value of k equal to 5 was used in cross-validation according to (Figure6).

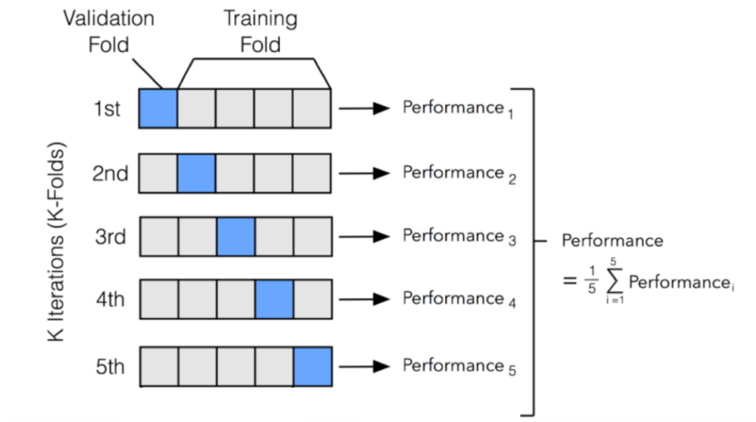


**Figure6.** Cross-validation, k-fold

(Figure7) shows the LSTM network architecture used in this research, which is based on data from the first to fourth periods. Input data has three dimensions: batch size, time period, and feature size. On top of the input nodes, an LSTM layer is placed on the network that models the time dependence between time periods. LSTM layer outputs also have three dimensions: batch size, number of time periods, and output node size. A dropout layer is then piled on top of the LSTM layer to cover the output by removing 50% of it. The dropout layer is used as a tuning technique to prevent overfitting. Since the shape of the LSTM outputs does not match the shape of the target, a dense layer with the sigmoid activation function is added to make the final predictions. The dense layer changes the shape of the LSTM output from three to two dimensions. In this way, the final outputs of the model correspond to the objectives in terms of data format. Adam's algorithm [12] is used to optimize the LSTM network and the epoch size of the 100 for training and binary cross entropy as a cost function. According to previous recommendations [13], one or two hidden layers are usually sufficient for neural networks to classify samples. So a hidden layer is selected.

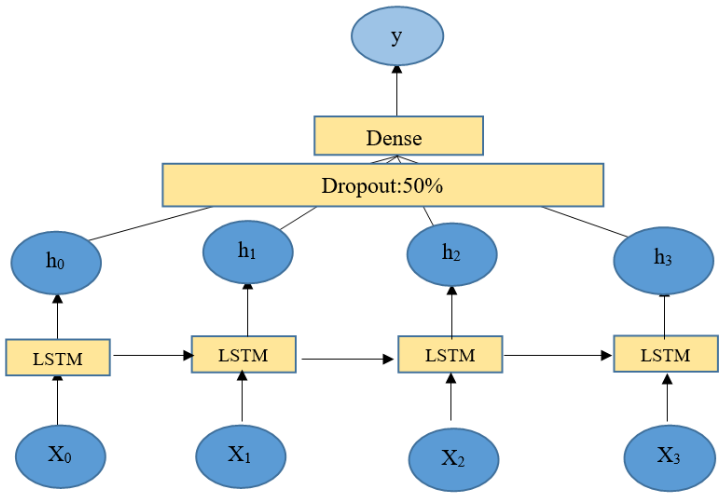


**Figure7.** Modeling time series behaviors with LSTM networks

**Analysis and evaluation**

As shown in (Figure8), LSTM network prediction improves over time, LSTM network gradually depicts improvement in academic performance prediction. Accuracy varies from the first to the fourth period, which indicates an increase in academic performance. The LSTM network predicted each student's academic performance with 83.14% accuracy in the first period, while 85.1% in the second period, 88.01% in the third period, and 94.02% in the fourth period. Therefore, the solution of this research is a vital determining factor in the early prediction of students at risk. Compares accuracy and loss values in 4 periods for LSTM network. It can be seen that the accuracy during the periods increases and the loss values decrement with increasing the periods, which indicates the strength of the model.

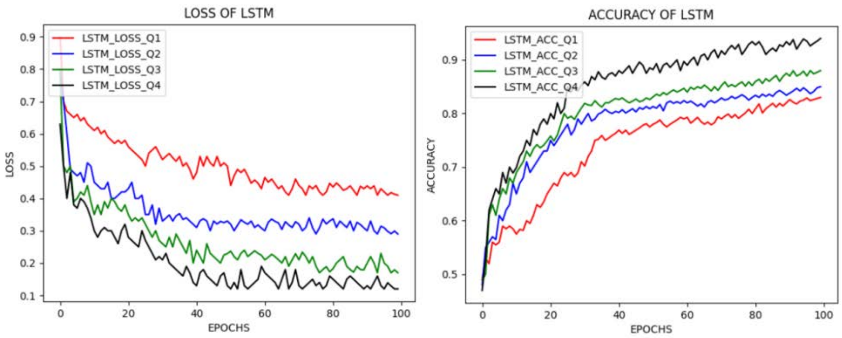


**Figure8.** Comparison of accuracy and loss in 4 LSTM network periods

The criteria for accuracy, precision, Recall, and specificity are shown in (Table3) and (Table4) for the LSTM network and the support vector algorithm.

Precision values in the LSTM network improved from 69.12% in the first period to 92.71% in the fourth period. Specificity values also improved from 88% in the first period to 96% in the fourth period. Similarly, recall values improved from 66.02% obtained in the first period to 90.09% in the fourth period.

Precision values in the support vector machine improved from 61.37% in the first period to 87.86% in the fourth period. Specificity values also improved from 88% in the first period to 96% in the fourth period. Similarly, recall values improved from 57.05% obtained in the first period to 79.73% in the fourth period.

**Table3.** Criteria for accuracy, precision, Recall and specificity in LSTM networks

| **LSTM** | First Period | Second Period | Third Period | Fourth Period |
|---|---|---|---|---|
| Accuracy | 83.14% | 85.1% | 88.01% | 94.02% |
| Recall | 66.02% | 77.50% | 80.18% | 90.09% |
| Precision | 69.12% | 78.28% | 85.42% | 92.71% |
| Specificity | 88% | 89% | 92% | 96% |

**Table4.** Criteria for accuracy, precision, Recall and specificity in SVM

| **SVM** | First Period | Second Period | Third Period | Fourth Period |
|---|---|---|---|---|
| Accuracy | 79.14% | 83.1% | 86.01% | 88.02% |
| Recall | 57.05% | 73.71% | 77.46% | 79.73% |
| Precision | 61.37% | 74.47% | 82.25% | 87.86% |
| Specificity | 86% | 87% | 90% | 93% |

As shown in (Figure9), the LSTM network performed significantly better than the support vector machine algorithm in predicting at-risk students. The LSTM network works well enough for sequential data. Hence, it analyzes student's behavior efficiently at periods and produces optimal results compared to the support vector machine algorithm.

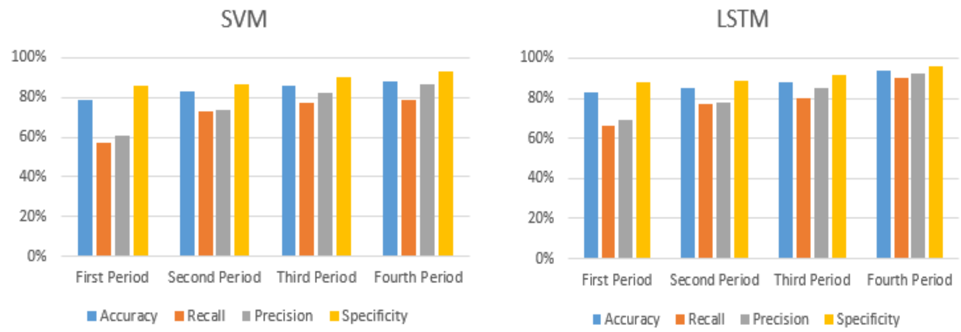


**Figure 9.** Evaluation with basic technique

Area Under the Curve (AUC) is another popular measurement criterion for algorithms. AUC is the area under the curve that receiver operating characteristic (ROC) [14]. Knowing that the higher the level below the AUC-ROC curve, the better the model has a chance of predicting Pass as Pass and At risk as At risk. (Figure10) shows the ROC curve for the two support vector machine and LSTM network used in this research. The AUC for LSTM and support vector machine was 0.936 and 0.882, respectively. The LSTM network performs better, so it is better to predict whether the student is at risk of falling.

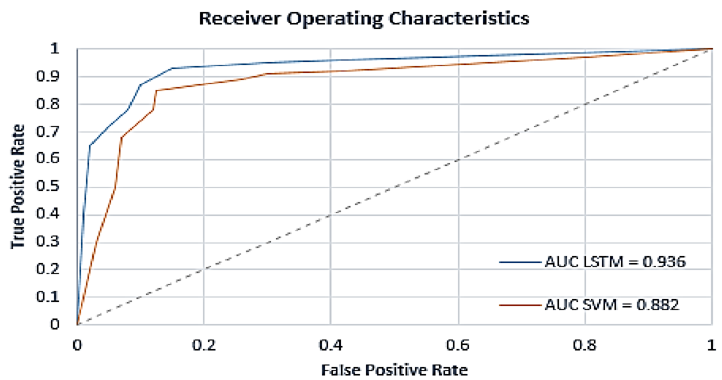


**Figure 10.** ROC curve

## Conclusion

Due to the shortcomings in the field of scientific supervision in universities, timely identification of the decrement in academic quality is essential. We have provided a high-precision, effective and efficient method for predicting and identifying students at risk of academic failure. First, a rich set of features have been collected, produced and calculated. One of the most important features that have been developed and used in this research is time-based feature. To separate the best feature set with maximum independence, we have used Pearson correlation, which ultimately led to the selection of 14 features. LSTM network and support vector machine algorithm have been used for modeling. We have also compared the LSTM network with the support vector machine algorithm in terms of predictive performance. The results have shown that the use of selected features and LSTM network predicted at-risk students with high accuracy. We have achieved 94% accuracy for the LSTM network and 88% accuracy for the support vector machine algorithm for predicting students' performances. The LSTM had more supervision over the students' course sequence pattern and activities, and has performed better than the SVM classifier, which relates to students' interactions. Such early forecasts help institutions provide timely intervention for at-risk students by providing a support system through counseling and alert emails. In addition, data-based analysis also helps decision-makers formulate student success policies based on their behavioral and interactive patterns. Criteria used to identify high-risk students help the educational community to record educational activities and corrective strategies for recording students' activities and behaviors to protect them through timely intervention.

## References


[1] Wasif, Muhammad, et al. (2019). Understanding student learning behavior and predicting their performance. Cognitive Computing in Technology-Enhanced Learning. IGI Global, 1-28.

[2] Costa, Evandro B., et al.( 2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior 73, 247-256.

[3] Yi, John C., et al. (2018). Predictive analytics approach to improve and sustain college student's non-cognitive skills and their educational outcome. Sustainability 10.11, 4012.

[4] Wilson, Joshua, et al. (2016). Comparing the accuracy of different scoring methods for identifying sixth graders at risk of failing a state writing assessment. Assessing Writing 27, 11-23.

[5] Marbouti, Farshid, Heidi A. Diefes-Dux, and Krishna Madhavan. (2016). Models for early prediction of at-risk students in a course using standards-based grading. Computers & Education 103,1-15.

[6] Aljohani, Naif Radi, Ayman Fayoumi, and Saeed-Ul Hassan. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. Sustainability 11.24, 7238.

[7] Osmanbegovic, E.; Suljic, M. (2012). Data mining approach for predicting student performance. Econ. Rev. J. Econ. Bus, 10, 3–12.

[8] Baradwaj, B.K.; Pal, S. (2012). Mining educational data to analyze students' performance. arXiv:1201.3417.

[9] Kovacic, Z. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. In Proceedings of the Informing Science and Information Technology Education Joint Conference, Cassino, Italy, 19–24 June.

[10] Watson, C.; Li, F.W.; Godwin, J.L. (2013). Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In Proceedings of the IEEE 13th International Conference on Advanced Learning Technologies, Beijing, China, 15–18 July; pp. 319–323.

[11] D. P. Kingma and J. L. Ba,( 2014). Adam: A Method for Stochastic Optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., Dec.
[12] S. Shankar and T. Purusothaman,( 2009). Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature Survey and Comparative Study, Int. J. Soft Comput. Appl., vol. 4, pp. 81–95.
[13] “Deep Learning - Ian Goodfellow, Yoshua Bengio, Aaron Courville - Google Books.” [Online].
[14] A. J. Bowers and X. Zhou, ( 2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes, https://doi.org/10.1080/10824669.2018.1523734, vol. 24, no. 1, pp. 20–46, Jan.